INSTITUTO SUPERIOR TÉCNICO
Universidade Técnica de Lisboa

Voice over IP System in an Academic Environment

André Filipe Abreu Regateiro

Dissertação para obtenção do Grau de Mestre em

Engenharia Informática e de Computadores

Júri

Presidente:	Prof. Luís Eduardo Teixeira Rodrigues
Orientador:	Prof. Fernando Henrique Corte Real Mira da Silva
Co-Orientador:	Prof. Teresa Maria Sá Ferreira Vazão Vasques
Vogal:	Prof. Paulo Rogério Barreiros D'Almeida Pereira

Setembro de 2007

Abstract

Voice over IP (VoIP) systems are assuming an increasing relevance in the telecommunications world, mainly for their potential to replace traditional voice systems and offer advanced applications.

This thesis reviews the main technological aspects of VoIP including the most relevant signaling, media and network protocols as well as the available software solutions.

It also documents in detail the VoIP system developed and deployed at IST and its integration with the remaining information systems of IST. This VoIP system described in this thesis is in production and is available for every IST user.

The operational and security implications that this new communication system introduces are analyzed, and applications created to interact with it are described.

Keywords: Voice over IP, SIP, NAT, PSTN Integration, VoIP security.

Resumo

Os sistemas de voz sobre IP estão a assumir um papel cada vez mais relevante no mundo das telecomunicações, principalmente pelo potencial para substituir os sistemas de voz tradicionais e pela possibilidade de oferecer aplicações avançadas.

Esta tese analisa os principais aspectos tecnológicos de Voz sobre IP incluindo os mais relevantes protocolos de rede, sinalizacao e media assim como as soluções de software existentes.

Também documenta detalhadamente o sistema de Voz sobre IP desenvolvido e instalado no IST, e a sua integração com os sistemas de informação existentes.

O sistema de Voz sobre IP descrito está em produção e disponível para todos os utilizadores do IST.

Analisa ainda as implicações operacionais e de segurança que este novo sistema de telecomunicações introduz, e descreve as aplicações criadas para interagir com o sistema.

Palavras Chave: Voice over IP, SIP, NAT, Integração com PSTN, Segurança em VoIP.

Acknowledgements

The author would like to thank the supervising professors Fernando Mira da SIlva and Teresa Vazão for their unwavering support and dedication to the project.

Thanks are also owed to some CIIST colleagues that helped during the project: Jorge Matias, Miguel Cabeça, Claudio Martins and Paulo Andrade.

Further thanks go to the Nucleo of Telecomunicações staff Victor Cóias and Paulo Rodrigues for their help with the PRI and BRI lines integration and to Luis Correia for the help in the configuration of the Alcatel PBX.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ALG Application Level Gateway

AGI Asterisk Gateway Interface

BRI Basic Rate Interface

DNS Domain Name System

ENUM Electronic Number Mapping

HTTP Hyper Text Transfer Protocol

ICE Interactive Connectivity Establishment

IETF Internet Engineering Task Force

ISDN Integrated Services Digital Network

ITU International Telecommunication Union

ITU-T ITU Telecommunication Standardization Sector

LDAP Lightweight Directory Access Protocol

MIKEY Multimedia Internet KEYing

MOS Mean Opinion Score

NAT Network Address Translation

POTS Plain Old Telephone System

PSTN Public Switched Telephone Network

PBX Private Branch Exchange

PRI Primary Rate Interface

QoS Quality of Service

RADIUS Remote Authentication Dial In User Service

RSVP Resource ReserVation Protocol

RTCP RTP Control Protocol

RTP Real-time Transport Protocol

SDP Session Description Protocol

SIP Session Initiation Protocol

SER SIP Express Router

STUN Simple Traversal of User Datagram Protocol Through NAT

SRTP Secure Real Time Protocol

TCP Transmission Control Protocol

TLS Transport Layer Security

TURN Traversal Using Relay NAT

UAS User Agent Server

UAC User Agent Client

UDP User Datagram Protocol

UPnP Universal Plug and Play

URI Uniform Resource Identifier

VoIP Voice over IP

VPN Virtual Private Network

Chapter 1

Introduction

Voice over IP, or simply VoIP, communications appeared in the last couple of years as the potential replacement for traditional voice communications. In its simplest form, VoIP can be explained as the digital encoding or modulation of voice and subsequent packaging and routing using IP packets. VoIP systems have associated with them a number of advantages not found in the traditional voice communication systems such as:

Lower cost of communications The usual flat rate price structure of the Internet means that VoIP calls may be charged independently of the duration of the call. Furthermore, there is no differentiation between long and short distance calling, as the geographic location is inconsequential.

Integration of data and voice infrastructure Using the Internet for voice communications means that a company only has to maintain one network and thereby reduce costs in equipment, infrastructure and personnel.

Better use of the bandwidth The packet switch approach of the Internet means that information of different users and applications can be transmitted at the same time maximizing the throughput.

Advanced Applications Due to the all IP approach of a VoIP system, the creation of new and advanced applications such as voice-mail, virtual call center or presence becomes easier and more flexible.

Tradicional voice communication systems, also known as Public Switched Telephone Network (PSTN) or Plain Old Telephone System (POTS), are circuit switched systems which require that for every call a logical circuit has to be created between both endpoints. This circuit has to be maintained for the entire duration of the call, whether the line is being used to transfer information or not.
VoIP systems, on the other hand, only create a packet when there is information to be transmitted. These packets are sent through a shared channel which can be used by others applications and users. This, used together with silence suppression technologies, allows for an extremely efficient use of the channel and a reduced call cost.

There are however some drawbacks in the use of VoIP systems, namely:

- The reliability of data networks is not as assured as the reliability of conventional voice networks, and downtimes of data networks are typically higher than those of voice ones.

- Data networks were build to transport large amounts of information and not for real-time communications, as such the latency[1] and jitter[2] issues of the Internet become real problems in the VoIP world.

The objective of this dissertation was to build a Voice over IP system at IST, integrate it with the existing voice system and provide the system to IST users. A complete VoIP system was created from the ground up, including the installation and configuration of several VoIP server components and its interconnection with the PSTN. Several applications were also created that interact with the system to provide advanced services to the users.

This document will present the system deployed to introduce a VoIP system at IST. It will also serve as future memory for the maintenance of the system.

It is organized as follows:

Section 2 will showcase the most significant technologies related to VoIP.

Section 3 will analyze the architecture of the VoIP solution deployed at IST and its most severe challenges.

Section 4 will focus on the security related issues surrounding the solution.

Section 5 will present some of the advanced applications developed with the system.

The main conclusions of this work are summarized in section 6.

[1]Time a packet takes to reach its destination.

[2]Variation of latency over time.

Chapter 2

Technology

VoIP technology can be thought of as an umbrella for a series of different technologies for different purposes. Among this technologies are signaling protocols, media transmission protocols and network protocols. In this section, the most important protocols relating to VoIP are analyzed. We will take a bottom up approach, presenting first the lower levels of the technology and showing how they are integrated with the upper levels.

2.1 Codecs

Codecs, short for coder/decoder, are algorithms for converting and compressing multimedia data into a digital signal in order to transport it over the network.
Codecs are the base of all VoIP technology, in the sense that they are the ones responsible for the translation between the analog and digital world. The voice codecs used for VoIP differ greatly in the quality of audio they provide and the bandwidth and processing time they require.
VoIP phones typically have to chose a trade-off level between audio quality, bandwidth and processing time, but with the current high speed of network links and high CPU speeds, the higher quality codecs are most commonly used.
Audio quality can be measured through the Mean Opinion Score (MOS), a subjective numerical indicator created by the ITU Telecommunication Standardization Sector (ITU-T). The MOS score is obtained by averaging the results of a set of standard tests where a number of listeners rate the heard audio. The score is expressed as a single number from 1 to 5, where 1 is lowest perceived quality, and 5 is the highest perceived quality. Table 2.1 [Wal05] compares the characteristics of the most common voice codecs.

2.2 Real Time Protocol

The Real-time Transport Protocol (RTP) defines a standardized packet format for delivering audio and video over the Internet in a near real time fashion.
It was developed by the Audio-Video Transport Working Group of the Internet Engineering Task Force (IETF) and is currently specified by RFC 3550 [HS03].

Codec	Bandwidth	Packet interval	Processing Intensity	MOS
Linear	128 Kbps	20ms	Low	4.5
G.711	64 Kbps	20ms	Low	4.1
G.726	32 Kbps	20ms	Medium	3.8
G.728	16 Kbps	10ms	High	3.6
G.729	8 Kbps	10ms	High	3.7
G.723	6.3 Kbps	10ms	High	3.6
GSM	13 Kbps	20ms	High	3.5

Table 2.1: Audio Codec Characteristics

RTP does not have a standardized Transmission Control Protocol (TCP) or User Datagram Protocol (UDP) port assigned but it is generally configured to use UDP ports 16384 to 32767.
In the typical VoIP network RTP is only responsible for the transmission of media. It does so by dividing the encoded audio informations in packets and sending then thought the Internet. It also adds control information such as playing order and session identification.

The bandwidth indicated by the codecs in the previous section refers only to voice information. To each audio packet it has to be added the RTP header, the UDP header, the IP header and the Ethernet or other OSI level 2 header. A typical G.711 packet contains 1280 bits of voice information regarding a 20 ms period and a minimum of 528 bits of headers, raising its required bandwidth to 90.4 kbps for a unidirectional VoIP session. The 528 bits of the header are composed of 12 bytes of the RTP header, 8 bytes of the UDP header, 20 bytes of the IP header and 26 bytes of the Ethernet header.

RTP is often accompanied by RTP Control Protocol (RTCP) to provide out-of-band control information for a RTP stream. Its job is to periodically send control packets to the participants indicating statistics such as bytes and packets sent and received, lost packets, jitter and round trip time. VoIP phones can use this information to change the codec choice during the session to better suite the network status. It can also be used to control and fine tune Quality of Service (QoS) settings in all the equipment involved in the media session transmission.

2.3 Session Initiation Protocol

The Session Initiation Protocol is an application layer protocol for creating, modifying and terminating media sessions between end-points. The protocol was created by the IETF and its latest specification is RFC 3261 [JR02a]. SIP supports five facets of establishing and terminating multimedia communications:

1. User location: determination of the end system to be used for communication.

2. User availability: determination of the willingness of the called party to engage in communications;

3. User capabilities: determination of the media and media parameters to be used;

4. Session setup: establishment of session parameters at both called and calling party;

5. Session management: including transfer and termination of sessions, modifying session parameters, and invoking services.

The protocol was designed to be simple, flexible, application and transport independent. Initially only implementing six methods that could be used on top of UDP or TCP.

The SIP protocol was heavily inspired by the Hyper Text Transfer Protocol (HTTP) protocol and shares with it several characteristics such as the Request Response paradigm and the numerical status codes. SIP, as HTTP, was also designed to have a low overhead being completely text-based and human-readable as can be seen in Fig. 2.1.

The port specified by the RFC is 5060, but it is usual to see implementations use the 5060-5070 UDP range.

The SIP for signaling and RTP for media combination was able to become the *de facto* standard for VoIP and has dethroned ITU-T H.323 umbrella protocol stack as the most used VoIP protocol.
H.323 has still a strong presence in the telecommunications service providers back-bone, but things seem to be changing there too. Other protocols like Megaco, SCCP or MGCP remain a niche market.
There are however very strong user bases in commercial VoIP systems such as Skype, but because of their proprietary protocol use they can not interoperate with other networks.

A SIP message is composed by several lines terminated with CR+LF. Each line specifies a different header field required for the session, Table 2.2 lists the most important headers. An optional body can be inserted in the message.

The protocol exchanges information via a series of Requests (messages from the client to the server) and Responses (messages from the server to the client).

2.3.1 SIP Requests

IETF has implemented the following SIP requests:

In RFC 3261 [JR02a]:

- INVITE, Indicates a client is being invited to participate in a call session.

```
INVITE sip:example@ist.utl.pt SIP/2.0
Via: SIP/2.0/UDP voip.ist.utl.pt;branch=t53G4bKas4asdhds
Max-Forwards: 70
To: Example <sip:example@ist.utl.pt>
From: Andre <sip:AndreRegateiro@ist.utl.pt>;tag4518321774
Call-ID: a84b4c76e66710@voip.ist.utl.pt
CSeq: 314159 INVITE
Contact: <sip:AndreRegateiro@voip.ist.utl.pt>
Content-Length: 80
```

Figure 2.1: A SIP Invite Example

Field	Use
Call-ID	Unique session/client identifier
Contact	Client identifier and locator
Content Length	Size in bytes of the message body
Content Type	Type of information contained in the body
From	Requester identifier
To	Receiver identifier
Cseq	Identifies a particular message in the session
Via	Identifies the next hop on the path

Table 2.2: Main SIP headers

- ACK, Confirms that the client has received a final response to an INVITE request.
- BYE, Terminates a call and can be sent by either the caller or the callee.
- CANCEL, Cancels any pending searches but does not terminate a call that has already been accepted.
- OPTIONS, Queries the capabilities of servers.
- REGISTER, Registers the address listed in the To header field with a SIP Registrar.

In RFC 3262 [JR02b]:

- PRACK, Provisional acknowledgment.

In RFC 3265 [Roa02]

- SUBSCRIBE, Subscribes for a Event of Notification from the Notifier.
- NOTIFY, Notify the subscriber of a new Event.

2.3.2 SIP Responses

The SIP protocol implements a three digit response code similar to HTTP:

1XX - Informational

2XX - Success

3XX - Redirection

4XX - Client Failure

5XX - Server Failure

6XX - Global Failure

2.3.3 SIP Nodes

SIP servers and endpoints are called nodes. They can be categorized as:

- SIP Registrar: a database server that collects, stores and disperses information about its clients location.
- SIP Proxy: a server that routes or redirects SIP Requests on behalf of one or more domains.
- SIP Outbound Proxy: a Proxy server that serves the task of connecting calls on behalf of a local network of SIP users.
- User Agent Client (UAC): responsible for sending SIP methods and receiving responses.
- User Agent Server (UAS): responsible for receiving SIP methods, processing them and return responses.

Most SIP phones implement both the UAC and UAS part of the protocol.

2.3.4 SIP Uniform Resource Identifier

SIP users can be referenced using a Uniform Resource Identifier (URI) that uniquely identifies a user in a particular domain, such as

`sip:AndreRegateiro@ist.utl.pt`

A SIP URI does not always correspond to a single phone. If a user is registered at more than one location then several phones can ring at the same time.

2.3.5 SIP Call Flow

A typical call flow starts with an Invite request to a proxy server as seen in Fig. 2.2. The proxy server remains in the signaling path but not in the media path.

2.4 Session Description Protocol

SDP is an IETF protocol rectified in RFC 4566 [MH06]. Its purpose is to describe the multimedia capabilities of the endpoints involved in a session.
SDP is, as SIP, a text based protocol and is usually packed inside the SIP Invite request and the SIP OK response. SDP indicates the host and port of the connection and the available capabilities using a series of fields listed in table 2.3.

Figure 2.3 shows part of a SDP payload of a SIP Invite request. This payload identifies the originator host and its available media capabilities, including in this case audio and video receiving capability. The audio and video streams are separated and each one is expected in a different port at the host. This port is specified in the 'm' field of the respective stream.

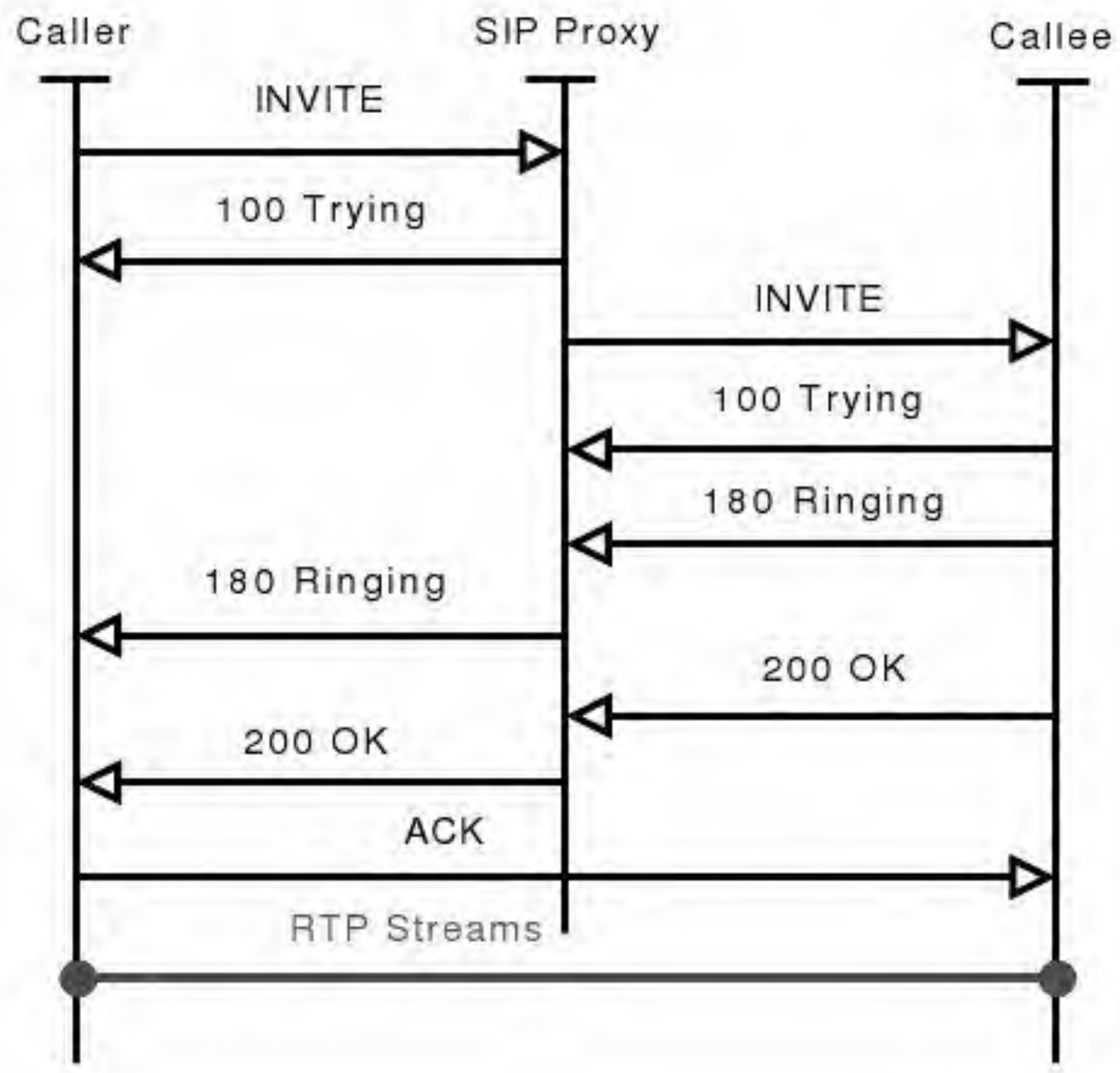

Figure 2.2: A typical SIP call setup flow

Field	Meaning
v	Protocol Version
o	Session Identifier and Originator
s	Session Name
i	Session Information
u	URI
e	Email Address
p	Phone Number
b	Zero or more bandwidth information lines
t	Time the session is active
m	Media name and transport address
a	Zero or more session attribute lines
i	Media title
c	Connection information

Table 2.3: SDP capabilities fields

```
v=0.
o=- 1170768338 1170768338 IN IP4 193.136.132.34.
c=IN IP4 193.136.132.34.
t=0 0.
m=audio 5000 RTP/AVP 101 96 3 107 110 0 8.
a=rtpmap:3 GSM/8000.
a=rtpmap:0 PCMU/8000.
a=rtpmap:8 PCMA/8000.
m=video 5002 RTP/AVP 31.
a=rtpmap:31 H261/90000.
```

Figure 2.3: Part of a SDP payload

2.5 Electronic Number Mapping

Being able to dial telephone calls the same way customers are used to is considered crucial for the convergence of PSTN and VoIP.
One simple approach is the Electronic Number Mapping (ENUM) [RD97] developed by the IETF that translates existing E.164 telephone numbers into SIP URI's using the Domain Name System (DNS) protocol and infrastructure.

In its simplest form ENUM enabled endpoints work by translating a phone number into a domain, the translation is performed by reversing the digit order and separating every digit with a dot, then the domain of the server[1] to be queried is added, for example:

```
+351 21 841 77 97 >> 7.9.7.7.1.4.8.1.2.1.5.3.e164.org
```

The resulting domain is then used to query a DNS server, the *Authoritative* server should have a NAPTR[2] record which resolves into a SIP URI. Several DNS servers already support ENUM. A configuration example for BIND can be seen in Fig 2.4.

```
$ORIGIN 7.9.7.7.1.4.8.1.2.1.5.3.e164.org
IN NAPTR 100 10 "u" "sip"  "!^.*$!sip:AndreRegateiro@ist.utl.pt!"     .
IN NAPTR 100 10 "u" "mail"  "!^.*$!mailto:AndreRegateiro@ist.utl.pt!" .
```

Figure 2.4: Example Bind configuration

The use of the DNS protocol is a particular happy choice since it is already massively deployed and it solves the technical problems of load balancing and geographic distribution.
Responsibility division however is an extremely difficult task, and there is no consensus on who should be responsible for maintaining the service, with telecommunications companies, regulatory agencies and end

[1] e164.org is a public ENUM directory.

[2] NAPTR stands for Naming Authority Pointer and is a newer type of DNS record that supports regular expression based rewriting.

users competing for control. Because of this, no single ENUM tree has emerged and several have to be checked while none can be considered authoritative. However, there is some convergence between international telecommunications agencies about the e164.arpa hierarchy as the base of a world wide ENUM system.

The Portuguese communications regulatory authority, ANACOM, has recently released the result of a public consultation for the deployment of a Portuguese ENUM tree. In this report ANACOM states that it will request the administration of the `1.5.3.e164.arpa` domain from the `e164.arpa` tree maintained by the International Telecommunication Union (ITU). The operational responsibility will be handed to the Fundação para a Computação Científica Nacional (FCCN). This should be the future unique and authoritative tree for the translation of Portuguese E.164 numbers.
As soon as this tree is available IST should require the administration of the `7.1.4.8.1.2.1.5.3.e164.arpa`, `8.1.4.8.1.2.1.5.3.e164.arpa` and `9.1.4.8.1.2.1.5.3.e164.arpa` domains.

2.6 Quality of Service

In the field of packet-switched networks the term QoS refers to control mechanisms that can provide different priorities to different data flows, or guarantee a certain level of performance to a data flow in accordance with requests from the application program. Quality of Service guarantees are important if the network capacity is limited, especially for real-time streaming multimedia applications such as VoIP which is very sensitive to jitter and delay.
Typical IP networks are provisioned to be *Best effort* and, therefore, do not guarantee timely or ordered delivery. This together with network congestion may hinder VoIP useless.
There are several approaches to provide QoS and they differ in efficiency and difficulty of implementation. The two most commonly used are Integrated Services and Differentiated Services.

2.6.1 Integrated Services

The idea behind Integrated Services or IntServ [RB94] is that every router in the system implements it, and every application that requires some kind of guarantees has to make an individual reservation. After that reservation is granted, and depending on the type of request, the network equipment guarantees the necessary bandwidth for the media session and limits the delay to a given amount. IntServ typically uses Resource ReserVation Protocol (RSVP) [RB97] to signal the network equipment. RSVP capable endpoints send a PATH message to the receiver who then replies with a RESV message which traces the path back to the sender. The routers between the sender and listener decide if they can support the reservation being requested, and if they cannot then send a reject message to let the listener know about it. Otherwise, and once they accept the reservation they have to carry the traffic and enforce the policy. If nothing is heard for a certain length of time, then the reservation will time out and be cancelled.

The problem with IntServ is that many states must be stored in each router. As a result, IntServ works only on a small-scale and is not a very popular QoS implementation.

2.6.2 Differentiated Services

Differentiated Services or DiffServ [SB98] is a simple and scalable mechanism to provide QoS. DiffServ is a coarse-grained, class-based mechanism for traffic management in contrast with IntServ which is a fine-grained, flow-based mechanism. DiffServ can be used to provide low-latency, guaranteed service to sensitive network traffic such as VoIP while providing simple best-effort traffic guarantees to non time critical services such as web traffic or file transfers.

DiffServ operates on the principle of traffic classification, where each data packet is placed into a limited number of traffic classes. The trafic class is indicated by encoding a 6-bit Differentiated Services Code Point (DSCP) value into the Differentiated Services (DS) field of the IP packet header. DiffServ recommends a standardized set of traffic classes to make interoperability between different networks and different vendor's equipment simpler. Most Networks use the following classes:

- Best Effort
- Expedited Forwarding
- Assured Forwarding

Network traffic can be marked at the source with one of this classes and DiffServ-aware routers implement Per-Hop Behaviors (PHBs), according with that class of traffic.

One advantage of DiffServ over IntServ is that it is much easier to implement and is already present if not used in most network equipment. However, the way individual routers deal with the type of service field is somewhat arbitrary and so it is difficult to predict end-to-end behaviour. The main concern is that considering that DiffServ works by dropping packets selectively, traffic on the link in question must already be very close to saturation. For this reason, some think that DiffServ will always be inferior to adding sufficient network capacity to avoid packet loss on all classes of traffic.[3] This point extends to implementing QoS in general and it has plenty of arguments supporting it:

- It would require complex Service Level Agreements (SLA) involving several network peers and users.
- Implementation could require extensive all-or-nothing upgrades for network providers.
- It would be necessary to enforce SLA compliance and police network traffic.
- It could disrupt traditional network use in unforeseen ways.

Shalunov and Teitelbaum [BT02] present some further insight into this problems and its history as well as some possible solutions. The growing consensus about QoS provisioning of VoIP networks is that its not worth the trouble it might cause and should only be used as a last resort.

2.7 Integrated Services Digital Network

Integrated Services Digital Network (ISDN) [Uni93] is a circuit-switched telephone network system, designed to allow digital transmission of voice and data over ordinary telephone copper wires, resulting in better quality

[3] See Internet2 QoS working group conclusion [BT02] available at http://qbone.internet2.edu/

and higher speeds than that which is available with the PSTN system.
More broadly, ISDN is a set of protocols for establishing and breaking circuit switched connections, and for advanced call features.

There are two ISDN implementations:

- Basic Rate Interface (BRI) - consisting of 2 B (data) channels, each with bandwidth of 64 kbit/s, and one D (signaling) channel with a bandwidth of 16 kbit/s.

- Primary Rate Interface (PRI) - containing a greater number of B channels and a D channel with a bandwidth of 64 kbit/s. The number of B channels for PRI varies according to the nation: in North America and Japan it contains 23 B channels with an aggregate bit rate of 1.544 Mbit/s also know as T1; in Europe, India and Australia it contains 30 B channels with an aggregate bit rate of 2.048 Mbit/s also know as E1.

Although it was designed to do so, ISDN was never able to replace the traditional PSTN network. But PRI lines were able to become the standard in the high density, professional market.

Chapter 3

Architecture

3.1 SIP Express Router

SIP Express Router (SER)[ser] is an open source, high-performance SIP server, which can be configured to act as SIP registrar, proxy or redirect server. SER can be set-up to serve specialized purposes such as load balancing or SIP front-end to application servers.

SER's performance allows it to deal with operational burdens, such as broken network components, attacks, power-up reboots and a rapidly growing user population.

SER features complete support of RFC 3261 functionality, a variety of database backends, management features, multidomain hosting, ENUM, presence support, Remote Authentication Dial In User Service (RADIUS) or Database accounting and authorization and can be enhanced by a variety of additional SIP tools.

SER became the frontend of the IST's VoIP system, it is responsible for user registration and SIP routing for every application deployed.

3.2 Asterisk

Asterisk[ast] is a open source software implementation of a telephone Private Branch Exchange (PBX) , it allows a number of attached telephones to make calls to one another, and to connect to other telephone services including the PSTN.

The basic Asterisk software includes many features previously only available in proprietary PBX systems: conference calling, interactive voice response, and automatic call distribution. Asterisk also supports a wide range of VoIP protocols, including SIP, MGCP and H.323.

For interconnection with digital and analog telephony equipment, Asterisk supports a number of hardware devices, most notably single, double and quad span ISDN interfaces for interconnection to BRI or PRI lines and channel banks.

Asterisk is controlled by editing a series of configuration files. One of these, *extensions.conf*, is where the administrator defines what actions Asterisk will take when calls are answered. A native Asterisk language is

used to define contexts, extensions, and actions.

3.2.1 Asterisk Gateway Interface

Although Asterisk is already extremely flexible, it provides a standardized programing interface that can be used to extend its capabilities. This interface, called Asterisk Gateway Interface (AGI), provides access to the internal Asterisk states and functions to almost any programming language. It does so by allowing them to communicate through the standard input, standard output and standard error streams, much like the Common Gateway Interface (CGI) provided by most web servers.

Most of the advanced applications presented in section 5, as well as the authorization system of section 4, were developed using this interface.

3.3 PSTN Integration

One of the most important tasks of this work was the interconnection between the new VoIP system and the PSTN.
Presently the Alameda campus of IST uses an Alcatel Omni PCX 4400 PBX. This PBX serves 2700 phone extensions and is connected to the PSTN through 7 PRI lines providing 210 parallel voice channels. However 4 of this lines are configured to only originate calls, the remaining 3 being configured to only receive calls. Figure 3.1 showcases this situation.

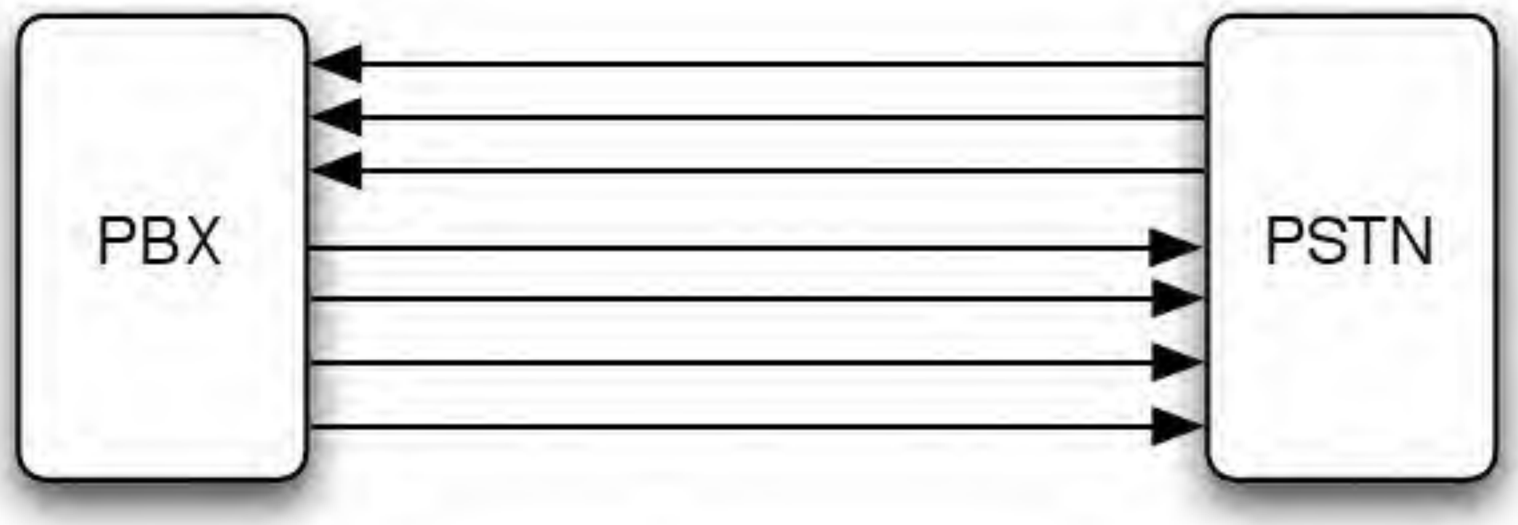

Figure 3.1: Original Phone System Configuration

The Alcatel PBX has an additional 4 BRI interfaces, three of them connected to the three Portuguese mobile phone operators; TNM, VODAFONE and Optimus. The fourth BRI interface was used for tests in the VoIP system.

The PRI interface cards in the PBX are quite expensive, and the PBX is not easily expanded to support an eight PRI card. So the solution found was to place the new VoIP system in the middle of an existing output primary line. This allowed the new system to communicate both with the Alcatel PBX and the PSTN directly while maintaining the exact same call capacity.

The PBX was reconfigured to accept calls coming from this interface, and to direct the unused 4000 - 4999 extensions to this line. The PSTN side suffered no alteration and so is still configured to only receive calls

through this line. All the calls that arrive at the voip-pri machine coming from the PBX and with destination onto the PSTN will be forwarded as is, so preserving the call load capacities of the previous voice system. All other calls that require use of the phone lines will be sorted in round robin to the available phone lines of the PRI interfaces.

The voip-pri.ist.utl.pt machine was equipped with a Digium Wildcard 205P with 2 E1 PRI interfaces, allowing it to connect to both sides simultaneously. The card is controlled with the Asterisk software.
The connection required the use of two 75/125 Ω BALUN convertors. These BALUN's (short for BALanced/UNbalanced) connectors are responsible for converting the 75 Ω balanced coaxial cable with a BNC connector of the PBX to the 125 Ω unbalanced cable with a RJ45 connector of the Digium card.

Both the lines were configured to use:

Swich Type - Euro ISDN

Framing - CCS

Coding - HDB3

Checksum - CRC4

Signaling - HDLC with FCS in channel 16

Side -Costumer Premise Equipment

This solution in showed in fig. 3.2.

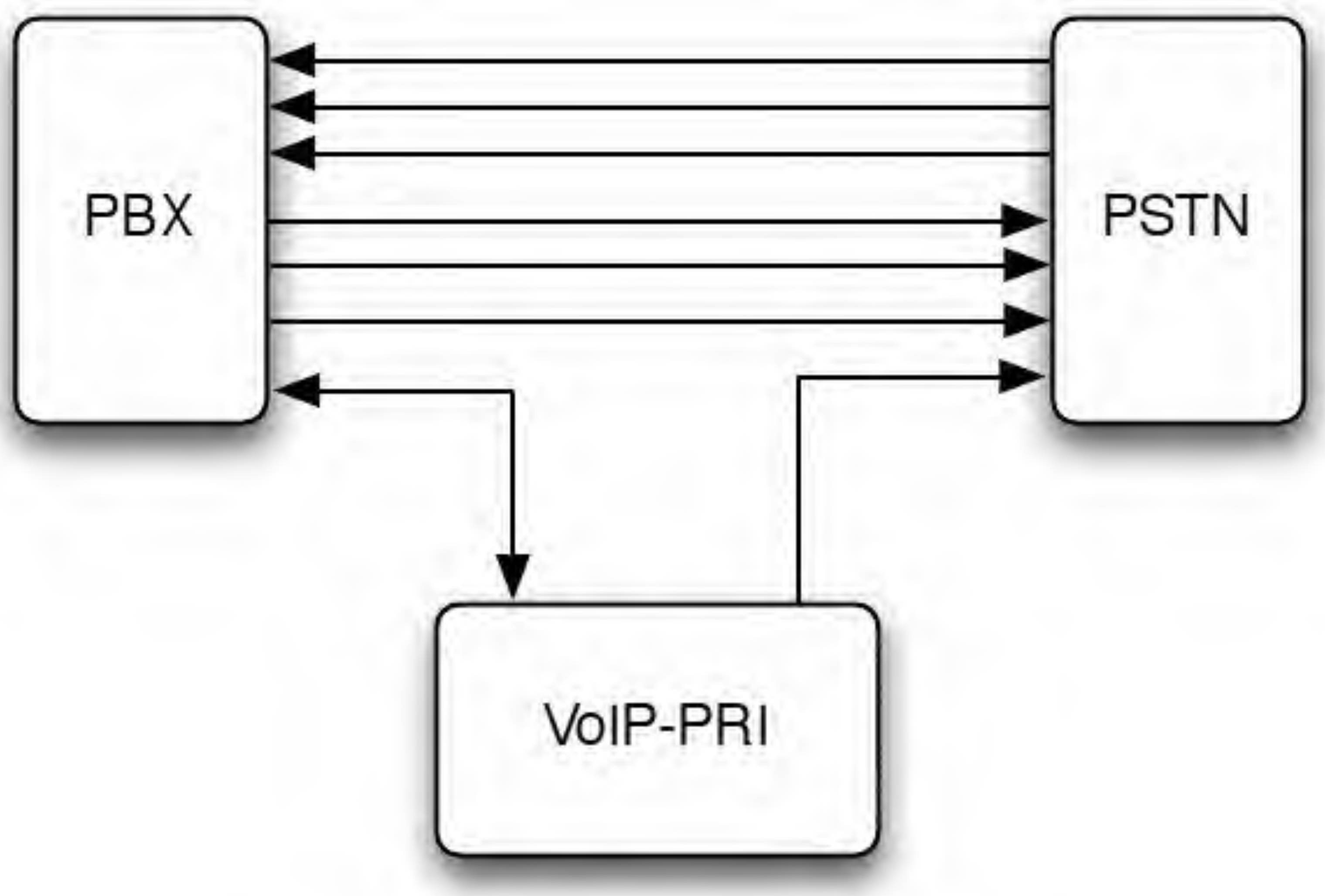

Figure 3.2: New Phone System Configuration

One important detail is the requirement of the telecommunications operator that every call with destination to the PSTN be appended with the prefix "1024". This is used in the operator side for call accounting. If a call is made without that prefix it will be rejected. Also the Caller ID field will be removed when calls are made to the PSTN, this is due to contract restrictions.

The 2700 phone extensions of the Alameda campus of IST were mapped to SIP URI's composed of <phone-extension>@ist.utl.pt. These are now publicly available SIP extensions that can be reached by any SIP user.

3.4 System Structure

For security reasons, the voip-pri.ist.utl.pt machine was isolated from the rest of the VoIP system, its only job is to interconnect with the PSTN.
For this reason it only runs the Asterisk software and an openssh-server to allow remote configuration.

The remaining system was installed in the voip.ist.utl.pt machine. This include:

- The SIP Express Router to register users and route SIP calls.
- The Asterisk PBX to support advanced applications.
- A Postgre SQL server database to support several applications and call accounting.
- The RTPProxy, STUN and TURN daemons introduced in the next section.
- A FreeRADIUS server for authentication.
- An Apache2 server to support the web applications.

This machine also features a BRI line connected to the PBX, this line was used during the development for tests, and is now configure to serve as backup for the PRI line. This BRI line is not allowed to make calls to the PSTN. The resulting system structure can be seen in fig. 3.3

3.5 NAT Handling

In computer networking, the process of NAT defined in RFC 3022 [PS01] involves re-writing the source and/or destination addresses of IP packets as they pass through a router or firewall. Most systems using NAT do so in order to enable multiple hosts on a private network to access the Internet using a single public IP address. According to IP specifications, routers should not act in this way, but many network administrators find NAT a convenient technique and use it widely.

In most configurations, NAT is used to translate between valid public IP addresses and private IP address. This private IP addresses are non-routable outside the local network were they are used and so they break the standard point-to-point communication of the Internet. This means that machines inside the network are able to communicate with each other and with machines in the outside through NAT. The machines on the exterior however cannot reach the machines inside. This is used by many administrators as a way to secure the local machines from intrusion. This implies that SIP clients behind a NAT cannot be reached directly from the exterior.

The way NAT works differs between implementations, but it can be summarized as follows. Each time a machine in the interior network sends a packet to the exterior, the NAT-box changes the source IP address

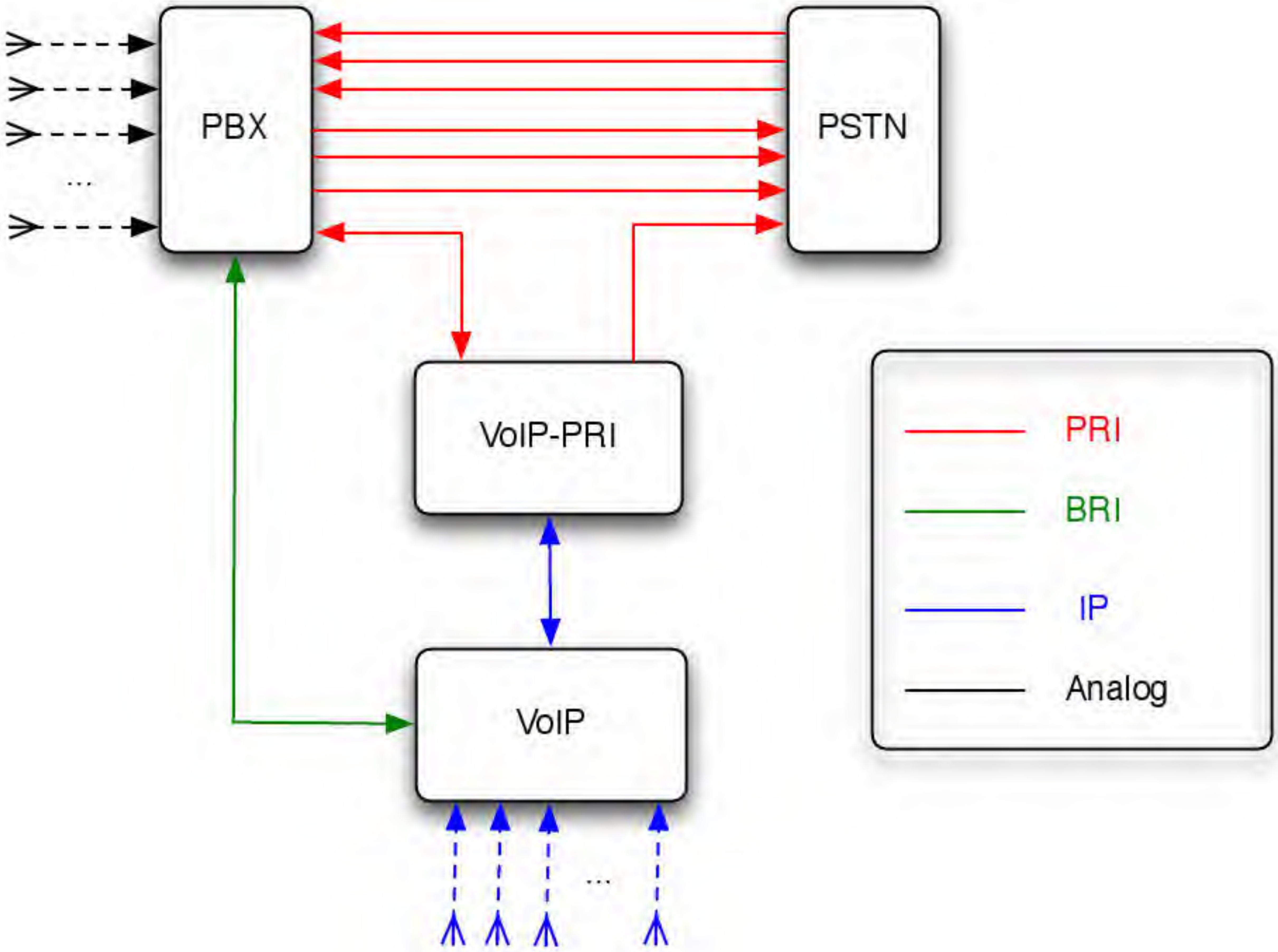

Figure 3.3: System Structure

and port of the packet from the private IP address and port of the sender machine to its own IP address and port.
This process is known as *source* NAT.
That packet is then forwarded to the Internet and the NAT-box creates a entry in a connection tracking (*conntrack*) table.
If the machine in the exterior wishes to reply it will do so and the packet will arrive at the IP address and port that the NAT-box specified. The NAT-box will then lookup this IP address and port combination and change the destination IP address and port of the original interior machine.
This process in known as *destination* NAT.

For TCP connections the NAT-box tracks the entire connection and only removes the entry from the conntrack table after the connection reaches the *closed* state. UDP on the other hand is a stateless protocol and so the entry will be removed from the table after a given timeout. In current Linux 2.6.* kernels, the timeout default is 30 seconds.
The NAT bindings are always created by inserting a tuple in the conntrack table. The way this tuple is built and used is what differentiates the type of NAT performed.
The tuples are created with a combination of the following fields:

- `Source IP` - The source IP address of the machine in the local network.
- `Source Port` - The source Port of the machine in the local network.
- `NAT IP` - The public IP address of the NAT-box.
- `NAT Port` - The source Port of the NAT-box.
- `Destination IP` - The IP of the machine being contacted in the exterior.
- `Destination Port` - The destination Port of the machine being contacted in the exterior.

The existing types of bindings are shown in table 3.1.

As a result of this differences the emerging behavior of the different types of NAT is different:

Full Cone NAT Each IP address and Port of the source machine is always mapped to the same IP address and Port of the NAT-box. This is why this type of NAT is also known as one-to-one NAT.

Full cone NAT `Source IP : Source Port : NAT IP : NAT Port`
Restricted cone NAT `Source IP : Source Port : NAT IP : NAT Port : Destination IP`
Port Restricted cone NAT `Source IP : Source Port : NAT IP : NAT Port : Destination IP : Destination Port`
Symmetric NAT `Source IP : Source Port : NAT IP & Port : Destination IP : Destination Port`

Table 3.1: Different NAT binding tuples

The first time the machine on the local network sends a packet to a exterior machine a mapping is created. This mapping can be used by any machine in the exterior to contact the local machine by simply sending a packet to the correct IP address and Port of the NAT-Box.

Restricted Cone NAT Each IP address and Port of the source machine is always mapped to the same IP address and Port of the NAT-box.
Unlike Full Cone NAT a machine on the exterior can only contact a local machine if it has received a packet from the local machine and so created a tuple in the NAT-Box in which the destination IP is the exterior machine IP.

Port Restricted Cone NAT This type of NAT is similar to Restricted Cone NAT but it includes the additional restriction that the exterior machine has received a packet in the Port indicated in the `Destination Port` field.

Symmetric NAT Each IP address and Port of the source machine and the IP address and Port of the destination machine is mapped to the same and unique NAT IP address and Port combination. This implies a new port is created in the NAT-box for every combination of `Source IP` : `Source Port` : `Destination IP` : `Destination Port` in use.

Appendix A showcases the behaviour of the several NAT types.
From a security standpoint the symmetric type NAT is the most restrictive and so considered the most secure. It is also the best performer of the 4 types in lookup operations to the table. This makes it the most popular implementation of NAT being used in several small office/home office Internet sharing devices as well as in Cisco equipment and in the Netfilter software of the Linux Kernel.

In every type of NAT it is impossible for a machine in the exterior to contact a machine in the interior. This problem is mitigated in SIP if every client is required to register in a SIP Registrar server before they are able to be reached. This register operation would create a binding in the NAT-box which would allow the client to be reachable from the server through the SIP protocol. The media session however would not be able to reach to client resulting in a call being established but no sound heard. The following subsections will analyze the main NAT solutions found.

3.5.1 Simple Traversal of User Datagram Protocol Through NAT

Simple Traversal of User Datagram Protocol Through NAT (STUN), as defined by RFC 3489 [JR03], is a network protocol which enables applications to discover the presence and type of NAT in their network. It also enables application to find out the public IP address of NAT-box.
STUN requires no changes in the NAT-box because it works in a client/server paradigm in which the client contacts the server indicating the IP address it is using. This server has to be reached in a valid and public IP address. The server compares that information with the source IP address and Port of the packet it received to determine if NAT is present, and if it is, the server then asks the client to perform a series of tests to determine the type of NAT. Finally the server indicates to the client the corresponding IP address and Port of the NAT-box.

With this information the client can, for example, indicate to the SIP server the NAT address where is is able to be reached, instead of its own private IP.

Therefore if any machine in the outside wishes to reach a machine in the inside, it should indicate that to the SIP Registrar server, which in turn relays that information to the inside machine through the binding it has in the NAT-box. The machine on the inside then connects to the remote machine indicating the IP address and Port of the NAT-box as its contact. This connection creates a new binding in the NAT-box, and if the NAT type is Full Cone, Restricted Cone or Port Restricted Cone it will use the same NAT IP address and Port combo that it used to contact the SIP Registrar server and communication will be possible. However if the NAT-box is of the Symmetric type it will create a new binding with a new NAT IP address and Port combo which the client will know nothing about and so communication will be impossible.

3.5.2 Traversal Using Relay NAT

The Traversal Using Relay NAT (TURN) protocol [JR05] appeared as a complement to STUN, its main job is to allow communication in the cases where STUN fails i.e. in the presence of Symmetric NAT. TURN accomplishes this task by positioning itself in the signaling and media path. The client indicates to the TURN server the destination its trying to reach and instead of sending the data to the destination directly it sends it to the TURN server which will *relay* it to the destination.

Both STUN and TURN require severe changes in the client's implementation and so it adoption has been weak. Companies such as SNOM and XTEN have adopted it.

3.5.3 Interactive Connectivity Establishment

The Interactive Connectivity Establishment (ICE) protocol is another protocol which enables the client to discover NAT presence and type. To do so it uses several other protocols, including STUN and TURN.
Based on the type of NAT found, the ICE protocol will try several solutions, and as a last resource might even supply a public IP address to the client. The ICE protocol is still a draft [Ros07] but is expected to be able to solve all of NAT problems. Its implementation however is not yet noteworthy.

3.5.4 Universal Plug and Play

The Universal Plug and Play (UPnP) protocol [For05] was created with the objective of allowing equipment to more easily self-configure and access the Internet. To do so it created an interface every UPnP enabled device must use to discover the network configuration and change it if necessary. In the case of NAT the client can ask the UPnP enabled NAT-box the type of NAT being used and can request the creation and tear-down of a specific binding.

The UPnP protocol effectively solves the NAT problem in the case where both the client and server are UPnP aware. Most small office/home office internet sharing equipment sold in the last couple of years are already UPnP aware, but several legacy ones are not. From the client side several softphones already implement it such as XTEN. The use of UPnP in the enterprise is considered a major security policy breach and its use will always be limited to the small office/home office devices.

3.5.5 Port Forwarding

A possible solution to reach a client behind NAT is to create static bindings in the NAT-box, the so called *Port Forwards.* If the client regularly uses the same IP address, signaling and media ports these rules can be inserted in the NAT-box and in the client configuration so that every traffic that reaches a set port will be relayed to the client. This method has obvious disadvantages: it requires control over the NAT-box, it can break communications for other users and so is only suitable to very small networks with few clients. Its use is only justified in small office/home office devices were UPnP is not available.

3.5.6 Application Level Gateway

A Application Level Gateway (ALG) is a add-on component to a NAT-box or firewall giving it specific application level knowledge about a protocol. In the SIP case a ALG is able to read and understand SIP packets and if necessary change them to allow communication.

In the source NAT phase instead of only changing the source IP address and Port it also changes the 'm' field of the SDP package indicating its own IP address and Port and automatically creating a NAT binding for this session. It does the reverse operation in the destination NAT phase. As it is able to understand SIP it can keep track of the session and can tear-down the NAT bindings when it sees a BYE request and acknowledge.

A ALG is a optimum solution for the SIP NAT transversal problem but it requires control over the NAT-box. Some commercial NAT-boxes and firewalls already support it. In the case of NAT-boxes implemented in Linux there exists a Netfilter patch which implements a SIP ALG.[200]

3.5.7 Media Proxy

One of the better solutions found to solve VoIP NAT problems is the use of a media proxy. This solution joins the best aspects of some of the previous proposals and uses some special characteristics of SIP to solve the problem for all NAT types.

The basic idea of the media proxy is to use a central point available at a public and reachable IP address that bridges the signaling and media paths of all end-points who are unreachable and only this ones. So each client connects to the media proxy an not directly to the other end-point. This connection can be done in a transparent way for the client thanks to the ability to change SIP packets transparently.

This method works as follows:

When a client contacts the SIP Registrar server with the Register request, the server compares the IP address of the SIP packet with the IP address of the IP header. If they differ, there must exist some form of NAT in the path. If this is case the SIP Registrar flags this on only this client as being behind NAT and saves its correct contact, i.e. the IP address and Port of the UDP packet.

To keep the binding on the NAT-box open the server periodically sends an empty Options request to the client.

When someone wishes to contact this client it sends a SIP Invite to ask the SIP Registrar for the contact, the server responds to this Invite request with an additional VIA header indicating all subsequent messages must pass through him.

When the Registrar receives the Invite for the client behind NAT it changes the 'm' field of the SDP packet

to refer the IP address and Port of the media proxy and then forwards the request to the client through the binding that was kept enabled by the periodical polling.
The client receives this request and if it accepts the call will respond with an OK response to the server indicating the IP and Port where it expects to be contacted. The SIP server will again change the SDP packet's 'm' field to indicate the media proxy IP address and port before forwarding the request.
The two clients will then try to reach each other by contacting the media proxy and by doing so will open a biding in the NAT-box. The media proxy will then simply relay all media traffic between the clients while the SIP server will be responsible for relaying the signaling information.

This method works even in the presence of Symmetric NAT because the clients open the media stream directly to what they think is the end-point and so create a valid biding in the firewall. Also both clients can be behind NAT-boxes and will still be able to talk to each other.

This method has several advantages over the previous ones including:

- It doesn't require any changes in the way the clients are implemented, they are totally unaware of any changes of behaviour.
- It doesn't require any changes in the NAT-box.
- It works for all types of NAT, even for the case where the two clients are behind NAT.

As a drawback this method requires the media path be extended to include the media proxy, this can cause a negative effect in the latency and jitter problem. It also doesn't scale very well, as all calls from clients behind NAT have to be relayed in a media proxy, this proxy can however be distributed by several machines mitigating the problem.

3.5.8 Discussion

There is no single perfect solution to address the NAT problem, but most of the problems can be solved by the following rules:

1. Avoid using NAT altogether, the simplest way to deal with NAT is to remove it from the equation. The benefits of NAT can be achieved with simpler firewall schemes and the IP addresses shortage should be addressed by introducing IPv6.
2. Add support to SIP ALG in every controlled NAT-box.
3. In the IST VoIP system a STUN and TURN server were introduced, some clients are able to solve their NAT problems this way.
4. For all other clients a media proxy scheme was implemented changing the SER routing behavior and adding a RTPProxy server.

Chapter 4

Security

The switch to a VoIP system presents several challenges to the security of communication systems. Traditional telephonic systems are usually proprietary and provide an isolated operational environment. VoIP systems, on the other hand, are vulnerable to the same threats that are encountered in typical Internet applications. Additionally the new applications supported by these systems present security issues that were not encountered in traditional voice systems.

4.1 Authentication

User Authentication was never an issue in the PSTN world. Usually, if the user had physical access to the telephone he was authorized to use it. In a SIP world user authentication, user authorization and call accounting are essential.

SIP supports Authentication through the REGISTER command. A user may require registration simply as a guest and so not have to use any type of authentication. If this mode is disabled, as it is in the IST VoIP system, the user receives a 401 - Not Authorized response. The user then has to provide some form of Authorization. The flow is showed in figure 4.1.

The REGISTER command supports two types of Authentication. The first is plain text password witch is obviously very insecure. Any device listening on the network would be able to catch the username end password.

The second type is Digest, which is very similar to the Digest authentication system used in the HTTP protocol. This Authentication system protects the password by only transmitting MD5 hashes. The known attacks against MD5 do not affect its use in this particular protocol. The mechanism works by challenging the user by inserting a random nounce and realm in the "401 - Not Authorized" response and requiring the user to reply with a request filed formed of:

```
MD5 ( nounce : Realm : MD5 ( Username : Realm : Password ) )
```

The server can then do the exact same operation and compare the results. To do so the server must be able to access the clear text password, or have access to a previously MD5 hashed version of the Username, Realm and Password fields.

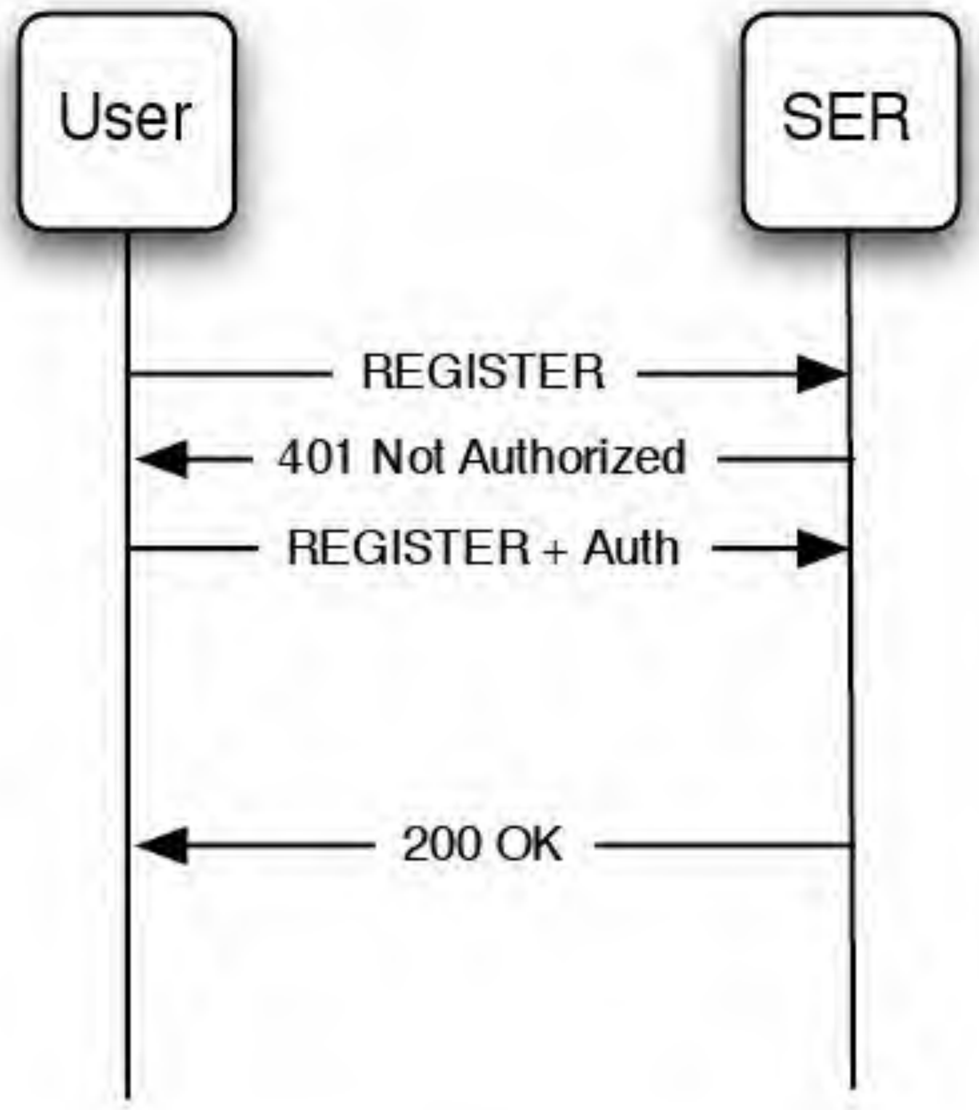

Figure 4.1: Simple SIP Registration

IST already provides a central authentication system that is able to authenticate through Kerberos, RADIUS and Lightweight Directory Access Protocol (LDAP) bind, but none of this methods is compatible with Digest. Kerberos does save the user passwords in plain text but accessing these fields would be a serious security violation.

This central authentication system provides users with an individual identifier know as ISTID. These IST IDentities were created to unify the several different identities of IST users.

A new Digest system could be created to support SIP, but that would require that the more than 10.000 IST users change their password or create a new one. Luckily there was already in place a HTTP proxy authentication mechanism which saves an MD5 hash of `Username : Realm : Password` in a LDAP field. SER does not support LDAP lookups, but it supports RADIUS authentication. A RADIUS server can then be used to authenticate the response against the LDAP hash.

The VoIP Authentication was then build to use exactly this method and is represented in figure 4.2. This method allowed to map every IST user to a SIP URI of the form sip:<ISTID>@ist.utl.pt.

The realm of the SIP Digest and HTTP digest differ, the SIP realm represents only the network, the proxy HTTP Digest however uses user and network information to form the realm. In this case the realm saved in the LDAP entry is "utilizadores@proxy.ist.utl.pt".

The users are then required to use this string as the realm. This presents a problem because the SIP server appends the user to the domain to form the URI, and is not expecting that the user presents himself as `ISTID@utilizadores@proxy.ist.utl.pt`. This required a small source code alteration in the SER software.

The Freeradius[rad] server used also had to be altered so that it could fetch the correct LDAP field, in this case `istPersonTokenHA1`. This field contains the realm followed by the hash in a format which the RADIUS server was not able to understand.

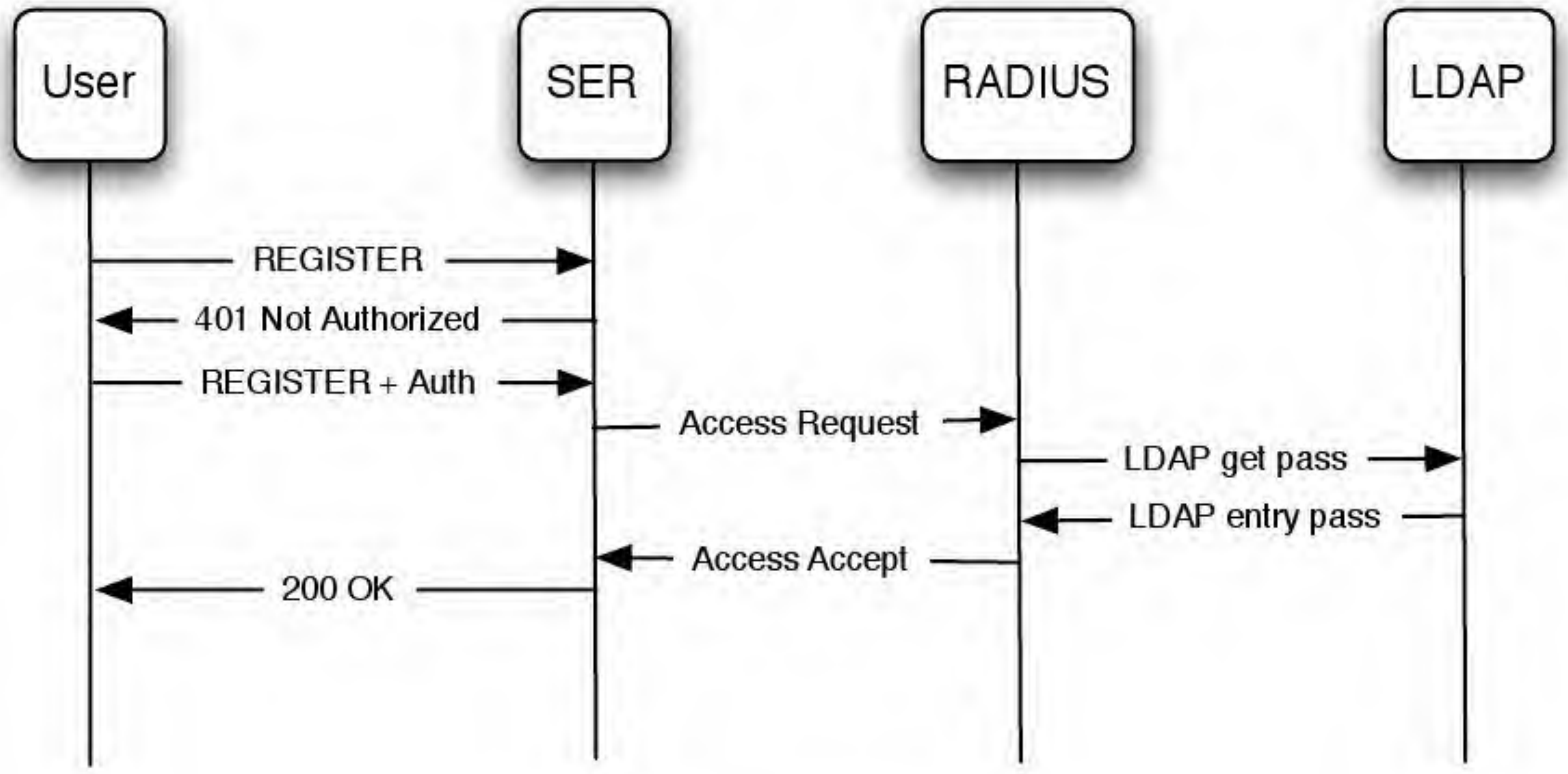

Figure 4.2: VoIP System Registration

These source code alterations are documented in the /root/SourceCodeAlterations file in the voip.ist.utl.pt server.

If the VoIP system user-base starts to grow it would be wise to install an openLDAP[lda] slapd daemon that mirrors the IST central LDAP servers in order to accelerate user authentication and reduce the load on the central authentication servers.

This system mandates that every IST user authenticates with his ISTID and does no allow the user to use a user name that is different from the authorization user name.
So every user is obligated to use the <ISTID>@ist.utl.pt SIP URI. However, after the user has successfully registered, a script is run to add that user email address as an alias to the SIP URI. The email address is retrieved from the central LDAP authentication servers and is only available to the users who have migrated to the new IST mail system. This allows every IST user to use the same contact for email and VoIP.

4.2 Authorization

The flexibility of a VoIP system allowed to create a much more fine grained authorization system. This system is able to control the type of destinations a user is capable of calling.
Four authorization categories were created:

- Free Calls - calls that incur no cost, like SIP calls and internal PSTN calls.
- National calls - calls with destination to continental Portuguese land lines.
- Mobile calls - calls with destination to any of the three Portuguese mobile operators.
- International - for all other destinations.

These categories were created basically for the different costs they incur. Every IST user is able to make free calls.

The responsibility for authorizing these calls is associated to the same individuals were responsible to do so in the tradicional communication system. In this system the IST faculty and staff are divided in financially independent groups called Cost Centers. Each one of this groups has one or more administrators that is responsible for maintaining a list of who is allowed to make calls from their phone extensions.

For the new VoIP system, a web interface[1] was created that allows these administrators to add and remove users from their cost center. Additionally, the Administrator can specify for each user the type of calls they are authorized to do.

Each user can only be associated with one cost center. If an administrator tries to add an existing user to his cost center, it will receive a warning containing the cost center were the user is currently registered. Figure 4.3 shows a screenshot of the Administrator Web Interface.

Since there is no central IT control of the administrators of these cost centers a second web interface[2] was created that allows the system superuser to view, add and remove Administrators. This system was initially populated with a static list of the assiduity administrators for the same groups, they should roughly correspond with the telecommunications administrators. This system will be in place until the information about administrators is maintained by either the Fenix system or the IST central LDAP servers.

The persistence data for this application is maintained in the Administrator and User tables of the aptly named Cost Center Call Control (CCCC) Database of the Postgre SQL server.

The Authorization system is enforced by an AGI program running at the voip-pri machine.

The Authentication and Authorization systems are provided by different services running in different machines. This could allow attackers to bypass the Authentication scheme and present themselves with false credentials to the Authorization system. To prevent this the VoIP system was configured to make the voip.ist.utl.pt machine the only interface with the outside. Every user must be authenticated in order to reach the voip-pri.ist.utl.pt machine. The voip-pri.ist.utl.pt machine is configured to reject every call that has not originated in the voip.ist.utl.pt machine.

4.3 Accounting

Effective call accounting is an integral part of the security of the VoIP system. It allows users and administrators to control the costs of the communications and easily check for anomalies. The IST VoIP system logs every call made into a Postgres SQL database. The users can then access a web interface[3] that details the origin, destination, call duration and estimated cost of every call made in the last year. The administrators can access the same information for every user under their control through the administrator web interface.

The costs of every call are estimated based on the reported per minute and first minute cost of call to the different destinations.

[1] http://voip.ist.utl.pt/administrator
[2] http://voip.ist.utl.pt/root
[3] http://voip.ist.utl.pt/accounting

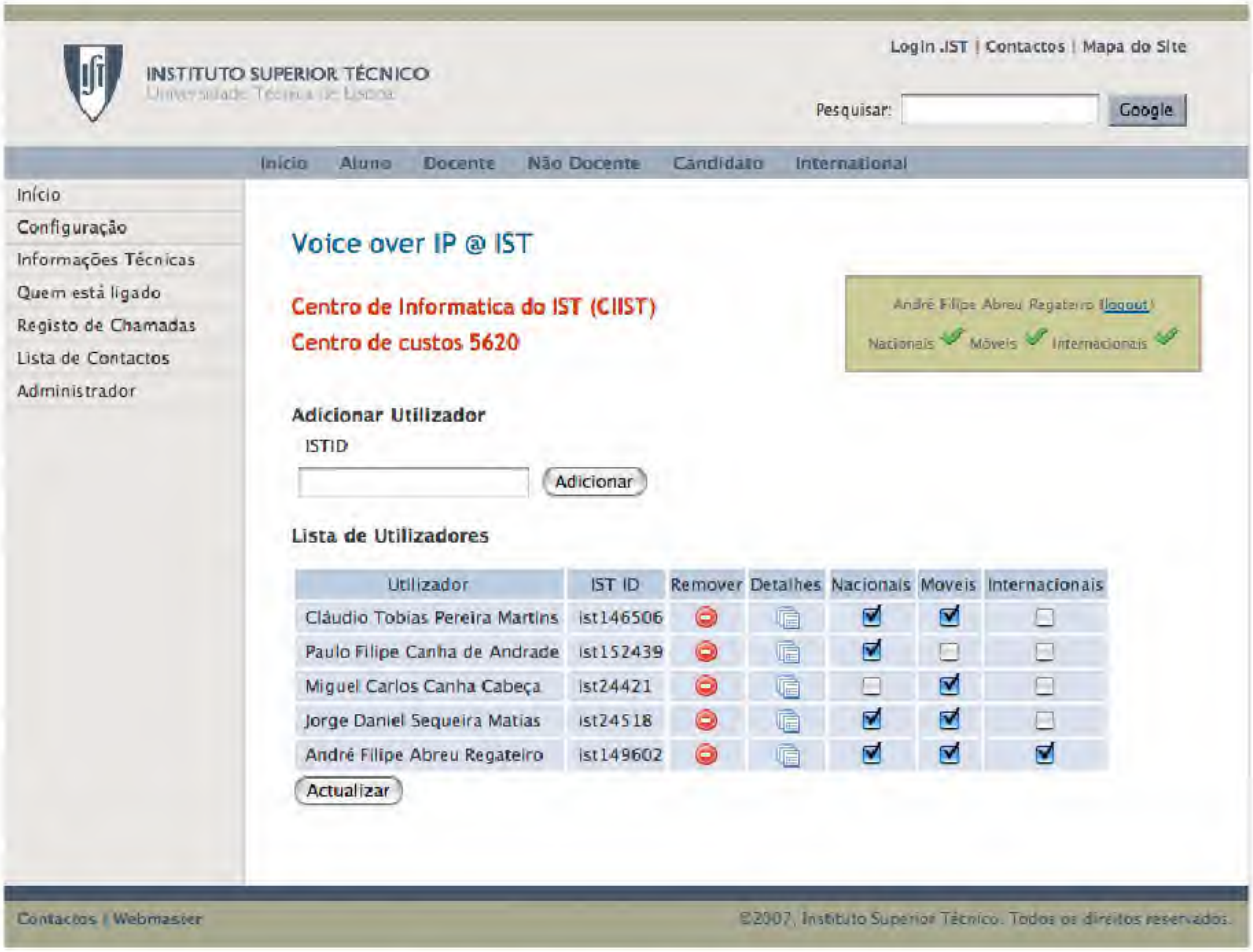

Figure 4.3: Administrator Web Interface

This web interface also allows the users to check the call authorizations they have been granted by the administrators. A screenshot of the Accounting interface is shown in figure 4.4.

The persistence data for this application is maintained in the Call Detail Record (CDR) of the Asterisk Database of the Postgre SQL server.

Both the Accounting and Authorization web interfaces authenticate users and administrators through the Central Authentication Service (CAS) protocol. CAS is a single sign-on protocol designed to allow web applications to authenticate users against a trusted central server. The CAS protocol involves a client web browser, the application requesting authentication, and the CAS server. When the client visits a protected application, it will be automatically redirected by the application to CAS. CAS will validate the client's user ID and password via a secure database, in this case the IST central LDAP servers. If the user ID and password are valid, CAS redirects the client to the application with a random number called a ticket. The application opens an HTTPS connection directly to CAS, and provides its own service identifier and the ticket. CAS then tells the application the user ID if the ticket is valid for that service identifier.

Figure 4.4: Accounting Web Interface

4.4 Confidentiality

Confidentiality refers to the need to keep information secure and private. In a IP telecommunications system eavesdropping on conversations is an obvious concern, but the confidentiality of other information of the call must be protected to defend against toll fraud, voice and data interception, and denial of service attacks. Network IP addresses, operating system type, telephone extension to IP address mappings, and communication protocols are all examples of information that, while not critical as individual pieces of data, can make an attacker's job easier.

With conventional telephones, eavesdropping usually requires either physical access to tap a line, or penetration of a switch. Attempting physical access increases the intruder's risk of being discovered, and conventional PBXs have fewer points of access than VoIP systems. With VoIP, opportunities for eavesdroppers increase dramatically, because of the many nodes in a packet network.

4.4.1 SIP

RFC 3261 [JR02a] mandates the use of Transport Layer Security (TLS) [TD06] for proxies, redirect servers, and registrars to protect SIP signaling. Using TLS for User Agents is recommended.
TLS is able to protect SIP signaling messages against loss of integrity, confidentiality and against replay. It provides integrated key-management with mutual authentication and secure key distribution.

One drawback of TLS use in SIP scenarios is the requirement of a reliable transport stack (TCP-based SIP signaling). TLS cannot be applied to UDP-based SIP signaling. Also the use of TLS disrupts the media proxy scheme because it prevents the proxy server from altering the SIP packets. However the main problem is the lack of support for TLS in most hard and soft-phones.

Just as secure HTTP is specified with the "https:", secure SIP is specified with a URI that begins with "sips:".

The IST VoIP system allows the use of any SIP phones that support TLS.

4.4.2 RTP

Without protection, RTP is insecure, as a telephone conversation over IP can easily be eavesdropped. Additionally, manipulation and replay of RTP data could lead to poor voice quality due to jamming of the audio/video stream. Modified RTCP data could even lead to an unauthorized change of negotiated quality of service and disrupt the processing of the RTP stream. The Secure Real Time Protocol (SRTP) is a profile of the RTP offering not only confidentiality, but also message authentication, and replay protection for the RTP traffic as well as RTCP.

SRTP was standardized at the IETF in the Audio-Video Transport working group. It was released as RFC 3711 [MB04]. SRTP is independent of a specific RTP stack implementation and of a specific key management standard, but Multimedia Internet Keying (MIKEY) has been designed to work with SRTP.

Advanced Encryption Standard (AES) in counter mode is the default algorithm, if encryption is desired. The key derivation function is AES in counter mode with a 128-bit master key from the key management.

SRTP provides increased security, achieved by

- Confidentiality for RTP as well as for RTCP by encryption of the respective payloads;
- Integrity for the entire RTP and RTCP packets, together with replay protection;
- The possibility to refresh the session keys periodically, which limits the amount of cipher text produced by a fixed key;
- An extensible framework that permits upgrading with new cryptographic algorithms;

- A secure session key derivation with a pseudo-random function at both ends;
- The usage of salting keys to protect against pre-computation attacks;
- Security for unicast and multicast RTP applications.

Because SRTP is defined as an RTP profile it may be used with existing multimedia standards. For SIP or more precisely SDP enhancements have been defined to transport the key management data necessary for SRTP. Thus, the combination of SRTP and MIKEY may be used to provide end-to-end encryption

4.4.3 MIKEY

SRTP uses a set of negotiated parameters from which session keys for encryption, authentication and integrity protection are derived. Multimedia Internet KEYing (MIKEY) describes a key management scheme that addresses real-time multimedia scenarios (e.g. SIP calls and SRTP sessions, streaming, unicast, groups, multicast) and is currently being standardized within the IETF's MSEC group.

The focus lies on the setup of a security association for secure multimedia sessions including key management and update, security policy data, etc., such that requirements in a heterogeneous environment are fulfilled. MIKEY also supports the negotiation of single and multiple crypto sessions. This is especially useful for the case where the key management is applied to SRTP, since here RTP and RTCP may to be secured independently. Deployment scenarios for MIKEY comprise peer-to-peer, simple one-to-many, and small-size interactive group scenarios. MIKEY supports the negotiation of cryptographic keys and security parameters (SP) for one or more security protocols. This results in the concept of crypto session bundles, which describe a collection of crypto sessions that may have a common Traffic Encryption Key (TEK) Generation Key (TGK) and belonging session security parameters.

MIKEY has some important properties:

- MIKEY can be implemented as an independent software library to be easily integrated in a multimedia communication protocol. It offers independency of a specific communication protocol (SIP, H.323, etc.)
- Establishment of key material within a 2-way handshake, therefore best suited for real-time multimedia scenarios
- There are four options for Key Distribution:
  - Preshared-key
  - Public-key encryption
  - Diffie-Hellman key exchange protected by public-key encryption
  - Diffie-Hellman key exchange protected with preshared-key and keyed hash functions
- Re-keying Support
- Multicast Support (one sender)

As with SIP TLS there is also a lack of support for SRTP and MIKEY in most phones.

4.5 Phone Provisioning

As with almost any new or rapidly growing technology security issues arise. This happens because security is almost never the main concern when developing new products and technology, and because new technologies present new challenges. The first wave of SIP hard-phones was particularly vulnerable to several types of attacks mostly of the DOS kind. As the technology has matured this problem has shrinked but is still present. Most system administrators are not comfortable exposing hard-phones to public Internet.

To protect them it is usual to isolate the hard-phones into separate VLAN's and use NAT to connect them to the outside. Besides being an operational hurdle, and as seen in section 3.5 this introduces several other problems. Another typical alternative is the use some sort of Virtual Private Network (VPN) technology that some phones support. VPN support in SIP phones usually comes with one of these tree types:

- Layer 2 Tunneling Protocol (L2TP);
- Microsoft's Point-to-Point Tunneling Protocol (PPTP);
- IPsec (IP security)

The first two types are becoming obsolete. The third one, IPsec should be the preferred form of VPN tunneling across the Internet.

4.5.1 IPsec

There are two basic protocols defined in IPsec [SK05]: Encapsulating Security Payload (ESP) and Authentication Header (AH). Both schemes provide connectionless integrity, source authentication, and an anti-replay service. The tradeoff between ESP and AH is the increased latency in the encryption and decryption of data in ESP and a "narrower" authentication in ESP, which normally does not protect the IP header "outside" the ESP header. Both schemes insert an IPsec header into the packet. IPsec also supports two modes of delivery: Transport and Tunnel.

Transport mode encrypts the payload and upper layer headers in the IP packet. The IP header and the new IPsec header are left in plain sight. So if an attacker were to intercept an IPsec packet in transport mode, they could not determine what it contained; but they could tell where it was headed, allowing rudimentary traffic analysis. On a VoIP network this would equate to logging which parties were calling each other, when, and for how long.

Tunnel mode encrypts the entire IP datagram and places it in a new IP Packet. Both the payload and the IP header are encrypted. The IPsec header and the new IP Header for this encapsulating packet are the only information left in the clear. Usually each tunnel is between two network elements such as a router or a gateway. In some cases, such as for mobile users, the tunnel could be between a router/gateway on one end and a client on the other end. The IP addresses of these nodes are used as the unencrypted IP address at each hop. Hence, at no point is a plain IP header sent out containing both the source and destination IP. Thus if an attacker were to intercept such packets, they would be unable to discern the packet contents or the origin and destination. Note that some traffic analysis is possible even in tunnel mode, because gateway addresses

are readable. If a gateway is used exclusively by a particular organization, an attacker can determine the identity of one or both communicating organizations from the gateway addresses.

IPsec allows nodes in the network to negotiate not only a security policy, which defines the security protocol and transport mode as described previously, but also a security association defining the encryption algorithm and algorithm key to be used.

The incorporation of IPsec into IPv6[Ege04] has increased the availability of encryption. VoIPsec (VoIP using IPsec) helps reduce the threat of man in the middle attacks, packet sniffers, and many types of voice traffic analysis. IPsec makes VoIP more secure than a standard phone line, where people generally assume the need for physical access to tap a phone line is deterrent enough.

IPsec and NAT compatibility is far from ideal. NAT traversal completely invalidates the purpose of AH because the source address of the machine behind the NAT is masked from the outside world. Thus, there is no way to authenticate the true sender of the data. The same reasoning demonstrates the inoperability of source authentication in ESP.

4.6 SPAM

Unsolicited commercial messages, better know as SPAM, is a huge inconvenient in the email system. In a VoIP system it could become an even bigger problem. Automated calling programs can easily be created to replay voice messages in a massive scale.

Due to the still small VoIP user base installed this has not yet become a problem, but its only a question of time.

Unfortunately almost no measures have been taken to mitigate this problem, mostly because there is no real practical solution. A possible one would be to use the presence extensions proposed in the RFC 3265 [Roa02] so that users would only accept calls from users they have previously marked as real contacts. This would still leave the problem of the first contact which would require some out of band communication. Some other solutions include forming web of trust schemes or disabling free calls.

Chapter 5

Advanced Applications

Switching to a VoIP system presents several advantages, but the biggest one by far is the ability to provide advanced application that were not possible or practical with the traditional voice systems. This section presents the advanced applications that were put in place in the IST VoIP system.

5.1 Click To Dial

This application provides a convenient address book and automatic call origination system. The main idea behind it is to centralize the contacts for each user in one place, and allow the user to simply click the contact it wants to reach, the application will automatically originate the call.
For this to work, the system first calls the user through his contact, when the user accepts the call, the system will then call the contact the user wants to reach.

The destination contacts can be either SIP contacts or PSTN extensions. If the user wants to he can also change his own contact so that the calls are originated to that phone. For security reasons the users contact is limited to SIP contacts or Alameda PSTN extensions.

The same authorization system is in place for this calls, so this application will only place the calls the user has been granted by the administrator. These calls will also be accounted as regular user calls and will appear in the accounting interface.

In a typical usage example, any IST staff can open the Click to Dial webpage[1] and click the contact it wants to reach. The system will then originate a call between the user PSTN phone extension and the destination. This web interface also allows the user to add and remove contacts and change its own contact. A screenshot of this application can be seen in figure 5.1.

The persistence data for this application is maintained in the contacts and source tables of the Click To Dial (c2d) Database of the Postgre SQL server. This application also uses the CAS system to authenticate users.

Hopefully users will start using this applications to originate most calls, as an additional advantage this allows the VoIP system to check if there is a free call path to the destination through a ENUM protocol lookup

[1] http://voip.ist.utl.pt/c2d

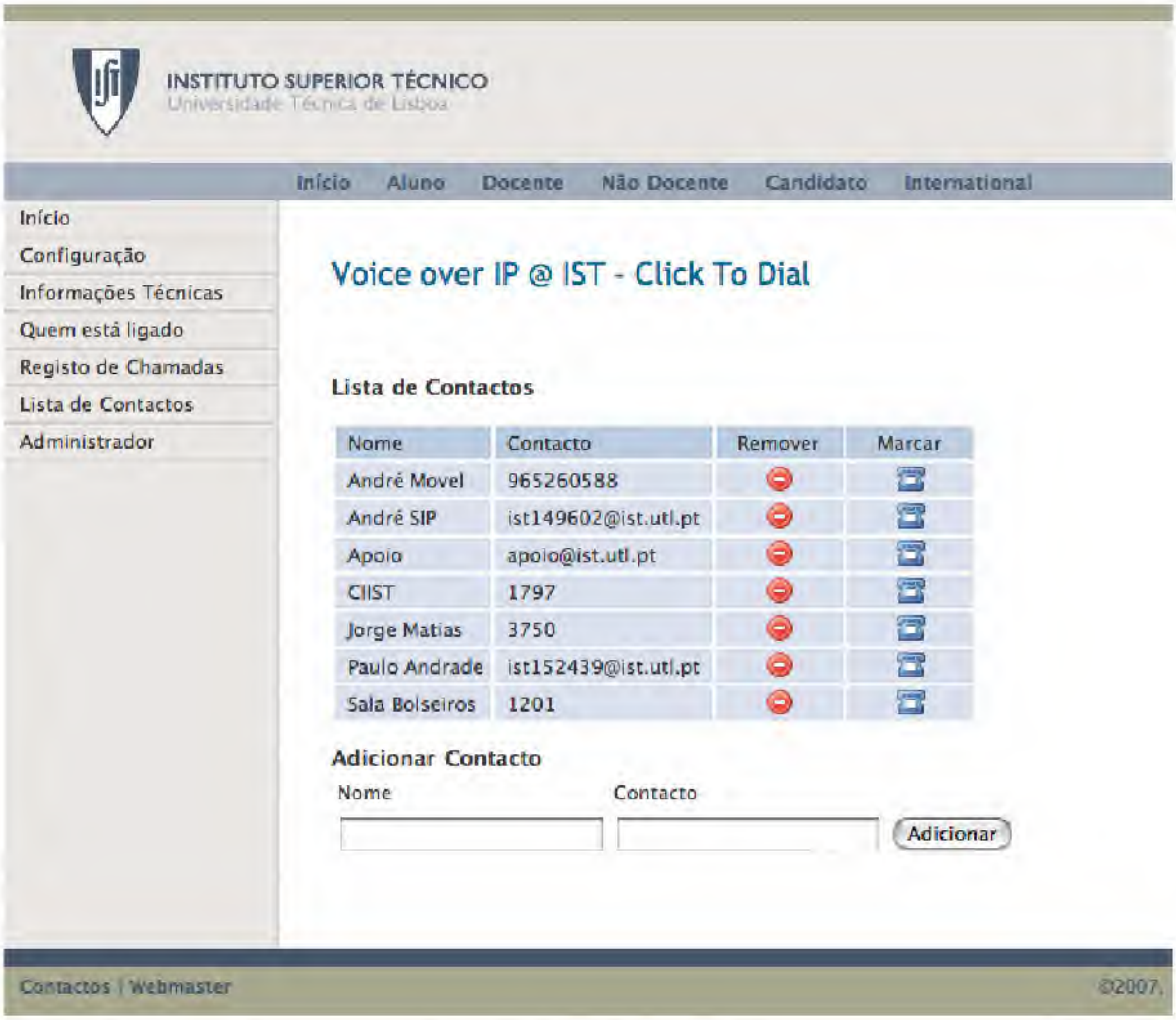

Figure 5.1: Click 2 Dial Web Page

for calls that otherwise would be delivered to the telecommunications operator and incur costs.

A drawback is that users will start forgetting phone numbers!

5.2 VoiceMail

This application is present in most commercial PBX systems, however it usually has a quite elevated cost. There are also several VoiceMail implementations for SIP calls, however none of them are appropriate to the current system, as they lack flexibility and/or secure access or require large and isolated storage space.

For the IST VoIP system a VoiceMail application for the SIP extensions was developed from the ground up using the existing IST resources. Instead of using separated storage space, this VoiceMail system uses the email accounts users have in the IST mail system.

It works as follows; When a call is made from anywhere to a unreachable IST user, the call is diverted to the VoiceMail system. A message is then played indicating that the user is not available and gives the caller the possibility to record a message. This message is them sent as an attachment in WAV format to the IST user email.

The voice message has a 30 second duration limit to avoid filling up users INBOX. The typical 30 second wav has an average size of 600 KB.

This application is quite simple and requires little to no maintenance, also it does not require separate storage space. As a drawback it does not allow the user to record his personal unavailable message.

This application can also be used to provide VoiceMail to the PSTN phone extensions but that would require support from the Alcatel PBX which is not present in its current firmware version.

5.3 Online Users

This simple web application application[2] shows which users are currently registered with the IST VoIP system. It shows every user that has registered with the IST VoIP system and what user agent they are using.

It is the only application that interfaces with the SER daemon. It does so through the only interface this server supports, a unix socket that accepts standard requests for info such as who is registered and which user agent is being used. It is also the only public web application and does not require authentication.

This application allows the world to see who is connected and what their contact is. This may become a problem for the unsolicited commercial messages that abound the email system. The SIP URI presented tough is only the one formed with the ISTID and has no connection with the IST users email.

If necessary the access to this application can be restricted to only authenticated users of the IST VoIP system.

[2]http://voip.ist.utl.pt/online

Figure 5.2: Online Web Page

5.4 PSTN to SIP Dialler

One of the challenges of the SIP and PSTN integration is the ability to call SIP extensions from a PSTN phone. This problem is mitigated for IST users by using the Click to Dial application to originate calls from their PSTN phone. But calling SIP users using only a PSTN phone is impossible. The main problem is the lack of an alphanumeric keyboard in the phones.

The ISTIDs used to form the URI all have the same format: ist[number] where [number] is composed of only digits and varies in size from 5 to 10 digits. These digits uniquely identify the IST user. An application was then created that uses this facts to bridge the SIP an PSTN worlds.

It works as follows:

1. The user calls a known phone number, 21 841 9832 from anywhere or 3832 from Alameda;

2. The application answers the call;

3. The user then uses the phone dial pad to insert the [number];

4. The application then bridges the call between the user and the sip:[number]@ist.utl.pt.

An example use can be seen in figure 5.3.

VoIP IST users are now able to disperse the 3832 + [number] or 21 841 9832 + [number] as their PSTN contact to their SIP phone. When calling from one of the Alameda PSTN phone extensions the user can dial the entire number all at once, the application will immediately pick up the call and start recording digits. When calling from any-other location the user has to wait for the 21 841 9832 phone extension to pick up and only then dial the ISTID digits. This extension beeps when it answers so that the user knows when to start dialing.

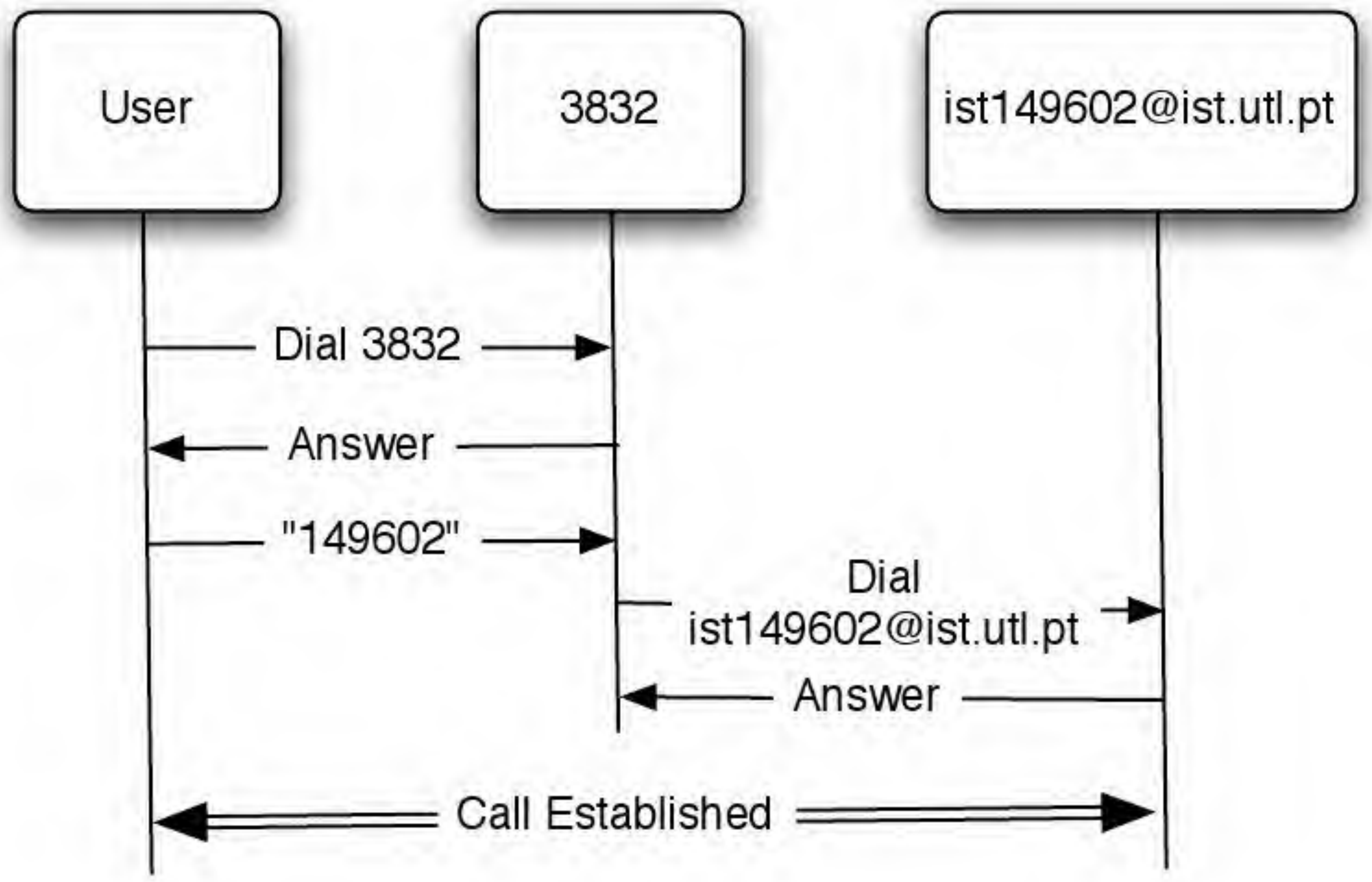

Figure 5.3: Example PSTN Dialler

5.5 Voice Dialer

This application was developed as a proof of concept to the use of a voice recognition system that allowed users to use voice do dial calls. The voice recognition system used was the Audimus voice recognition software from the *Laboratório de Sistemas de Lingua Falada* of Inesc-ID. This systems allows the voice recognition through the use of previously recorded voice models.

This application was implemented with both communication systems, VoIP and PSTN.

5.5.1 Architecture

The Audimus system functions in a client server architecture. The server has the voice models and accepts calls from the clients for recognition. The clients open an unidirectional stream to the server to send the audio, this audio is in the PCM format with an 8000 per second sample rate and signed 16 bits per sample. The server responds with a series of events like Start-of-Speech, End-of-Speech and Result.

Audimus provides a Java API that allows the construction of client applications. A client application was then created that interfaces with Audimus and with Asterisk through the AGI. The Asterisk server was configured to answer the 4000 PSTN extension and the 4000@ist.utl.pt SIP extension and then hand control to the program. The recognition model used has for a base 12 names of CIIST staff.

5.5.2 Develop Application

This application was developed as a daemon that waits a connection from the AGI interface. When a user dialed, the application instructed Asterisk to record the audio to a RAW format file. This recording had a 5

second limit and a silence timeout of 1 second. This means the user had 5 seconds to say the contact it wants to reach but does not have to wait the full 5 seconds.

The application would then use the sound eXange (sox) library to convert the RAW audio file to the 8000Hz 16bit signed format that Audimus expects. The result was then injected into a stream to the Audimus server.

When a recognition event was received the application translates the returned text into a contact. This could be done with the exact same contact database of the Click to Dial application. The application would then bridge the existing call with the required contact. An example call can be seen in figure 5.4.

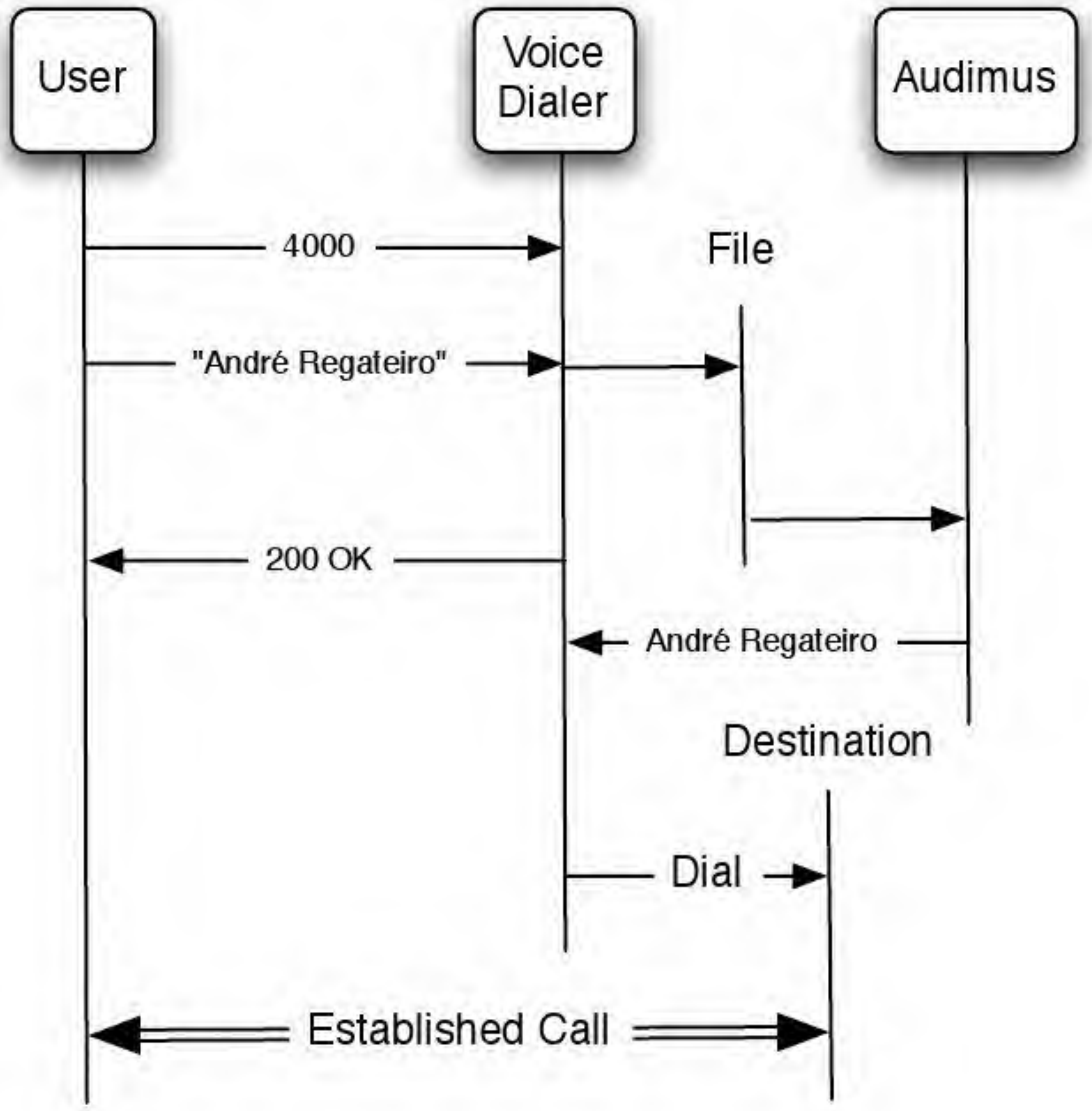

Figure 5.4: Example Voice Dialer

5.5.3 Results

For the 12 names model used the application has a very high accuracy, being able to dial the correct contact 90% of the calls. The wrong number was dialed 2% of the time, in the remaing 8% no call was establihed. There were no significant diferences between the SIP and PSTN results. Further tests would be required to test the acurracy of a voice model with more than 12 names.

This application was the only one developed that was not publicly released. To do so would require that a voice model was created for each user, and the ability to dynamically load and unload voice models. The last version of Audimus already allows this control over voice models, but the model generator is not yet matured.

Chapter 6

Conclusion

6.1 Discussion

VoIP technology is still somewhat immature and its implementation raises some concerns. However in the last couple of years it has evolved tremendously and it has steadily solved its major hurdles.
VoIP deployment should be evaluated primarily by its benefits in advanced application introduction and network integration.

The communications cost cutting effect of VoIP seems to be overrated and the complexity of the transition underrated. Nonetheless VoIP networks are the natural evolution of the PSTN network, and should replace it in the following years. The IST VoIP presents the first step in this path and was able to integrate both PSTN and VoIP networks.

The technological aspects of VoIP have matured and are ready for deployment. The NAT problem is not yet resolved but was mitigated and is now not a major concern.
The VoIP system integration with the existing IST services and PBX allowed the development of several new services for the IST staff.
The security of the VoIP system was one the guiding lines of the project and the system can be considered secure to a very comfortable level.
SPAM will probably become the main security concern.

6.2 Future work

Further work in this project is necessary, mainly for the integration of the IST VoIP system with the FCCN VoIP project.

The FCCN VoIP project aims to unify the Portuguese universities voice networks into a single VoIP network. This network will be able to make free calls within. The outbound calls would be delivered to one or more telecommunications company and the size of the network would enable it to achieve quite favorable commercial conditions. The IST VoIP system will then have to be adapted to this network.

Having this VoIP system in place allows the creation of several advanced telecommunication services that

can be specially fitted to the academic community.

Bibliography

[200] September 2007. Sip contract, http://people.netfilter.org/chentschel/docs/sip-conntrack-nat.html, available september 2007.

[ast] Asterisk, www.asterisk.org, available september 2007.

[BT02] Stanislav Shalunov Benjamin Teitelbaum. Why premium ip service has not deployed. Technical report, Internet2 QoS Working Group, 2002.

[DS05] Baruch Sterman David Schwartz. Nat traversal in sip, 2005.

[Ege04] G. Egeland. Introduction to ipsec in ipv6. Technical report, Eurescom, 2004.

[For05] UPnP Forum. Universal plug and play. Technical report, UPnP Forum, 2005.

[Geo04] Jeremy George. Sip.edu cookbook. Technical report, Internet 2, 2004.

[Goo02] B. Goode. Voice over internet protocol. Technical report, Proceedings of thee IEEE, 2002.

[GZH05] P. Ge Zhang; Hillenbrand, M.; Muller. Facilitating the interoperability among different voip protocols with voip web services. *Distributed Frameworks for Multimedia Applications*, 2005.

[HK06] James Philips Hechmi Khlifi, Jean-Charles Gregoire. Voip and nat/firewalls: Issues, traversal techniques, and a real-world solution. Technical report, Universite of Quebec, 2006.

[HS03] R Frederick H. Schulzrinne, S. Casner. A transport protocol for real-time applications. Technical report, IETF Request for Comments 3550, 2003.

[JR02a] G. Camarilho J. Rosenberg, H. Schulzrinne. Session initiation protocol. Technical report, IETF Request for Comments 3261, 2002.

[JR02b] H. Schulzrinne J. Rosenberg. Reliability of provisional responses in session initiation protocol. Technical report, IETF Request for Comments 3262, 2002.

[JR03] C. Huitema R. Mahy J. Rosenberg, J. Weinberger. Simple traversal of udp through nat. Technical report, IETF Request for Comments 3489, 2003.

[JR05] C. Huitema J. Rosenberg, R. Mahy. Traversal using relay nat. Technical report, IETF, 2005.

[JVM05] Leif Madsen Jim Van Meggelen, Jared Smith. *Asterisk, the future of Telephony*. O'Reilly Media, 2005.

[lda] Open ldap, http://www.openldap.org/, available september 2007.

[MB04] D. McGrew M. Baugher. The secure real-time transport protocol. Technical report, IETF Request for Comments 3711, 2004.

[MH06] C. Perkins M. Handley, V. Jacobson. Session description protocol. Technical report, IETF Request for Comments 4566, 2006.

[Net06] Newport Networks. Solving the firewall and nat traversal issues for moip, 2006.

[PS01] K. Egevang P. Srisuresh. Traditional ip network address translator. Technical report, IETF Request for Comments 3022, 2001.

[rad] Freeradius, http://www.freeradius.org, available september 2007.

[RB94] S. Shenker R. Braden, D. Clark. Integrated services in the internet architecture. Technical report, IETF Request for Comments 1633, 1994.

[RB97] S. Berson R. Braden, L. Zhang. Resource reservation protocol. Technical report, IETF Request for Comments 2205, 1997.

[RD97] M. Mealling R. Daniel. Resolution of uniform resource identifiers using the dns system. Technical report, IETF Request for Comments 1633, 1997.

[Roa02] A. B. Roach. Session initiation protocol specific event notification. Technical report, ETF Request for Comments 3265, 2002.

[Rob03] F. Robles. The voip dilemma. Technical report, SANS Institute, 2003.

[Ros07] J. Rosenberg. Interactive connectivity establishment. Technical report, IETF, 2007.

[SB98] M. Carlson S. Blake, D. Black. An architecture for differentiated services. Technical report, IETF Request for Comments 2475, 1998.

[ser] Sip express router, http://www.iptel.org/ser, available september 2007.

[SK05] K. Seo S. Kent. Security architecture for the internet protocol. Technical report, IETF Request for Comments 4301, 2005.

[TD06] E. Rescorla T. Dierks. The transport layer security protocol. Technical report, IETF Request for Comments 4346, 2006.

[Uni93] International Telecomunications Union. Integrated services digital network. Technical report, 1993.

[Wal05] Ted Wallingford. *Switching to VoIP*. O'Reilly Media, 2005.